Check for updates

Parallel deep neural networks for endoscopic OCT image segmentation

DAWEI LI,[1,4] JIMIN WU,[2,4] YUFAN HE,[2] XINWEN YAO,[1] WU YUAN,[1] DEFU CHEN,[1] HYEON-CHEOL PARK,[1] SHAOYONG YU,[3] JERRY L. PRINCE,[2] AND XINGDE LI[1,*]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA
[2]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
[3]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[4]Equal contribution
*xingde@jhu.edu

Abstract: We report parallel-trained deep neural networks for automated endoscopic OCT image segmentation feasible even with a limited training data set. These U-Net-based deep neural networks were trained using a modified dice loss function and manual segmentations of ultrahigh-resolution cross-sectional images collected by an 800 nm OCT endoscopic system. The method was tested on *in vivo* guinea pig esophagus images. Results showed its robust layer segmentation capability with a boundary error of 1.4 μm insensitive to lay topology disorders. To further illustrate its clinical potential, the method was applied to differentiating *in vivo* OCT esophagus images from an eosinophilic esophagitis (EOE) model and its control group, and the results clearly demonstrated quantitative changes in the top esophageal layers' thickness in the EOE model.

1. Introduction

Image-based medical diagnosis and prognosis rely on accurate interpretation of a large number of images. The task is laborious and prone to error. It would be highly beneficial to have a computer assisted system that can automatically detect and analyze abnormalities in the images and generate a more objective and quantitative analysis [1]. A crucial step toward such a system is to obtain accurate segmentation of features of interest in the images and to compute tissue characteristics such as shape, area, volume, thickness, and eccentricity [2].

Endoscopic OCT is an optical imaging modality capable of high-resolution, real-time, and three-dimensional (3D) imaging of internal luminal organs. It can be used in various clinical applications such as detection of subepithelial esophageal lesions [3] and coronary vulnerable plaques [4], where micro-structures yield important diagnostic information [5,6]. Automated segmentation of OCT images is a difficult task as segmentation algorithms are generally sensitive to speckle noise, intensity inhomogeneity, low image contrast, and other artifacts. So far, most efforts on endoscopic OCT image segmentation have focused on simple layer segmentation, such as stent strut detection and fibrous cap quantification of intravascular OCT images [7–10] and classification of esophagus OCT images [11]. Automated segmentation of OCT images with multiple layers has mainly targeted retinal and coronary images [12–15], again primarily using graph-based methods [14–18] or more recently deep learning techniques [19–27]. Endoscopic OCT images with multiple layers, however, often face elevated challenges such as complex layer boundary slopes due to tissue folding, blockage by mucus or some debris, and in-layer image intensity nonuniformity [28]. Recently, a graph-based method has been demonstrated for segmentation of multilayered endoscopic OCT esophagus images [29]. Compared to traditional image processing methods, deep learning techniques have higher tolerance to structure variability and require less prior

knowledge of the structures in the input images [30,31]. These advantages make them more attractive for developing automated segmentation methods [32].

In this paper, we proposed a robust segmentation method based on deep neural networks. To achieve robust layer segmentation, particularly with a limited training data set, U-Nets [19] were trained in parallel and then used to segment *in vivo* endoscopic OCT images. The paper first introduces the image collection process, describes the parallel training scheme in detail, and then demonstrates the performance of this method, including its robust segmentation ability, segmentation accuracy, and clinical potential. Robustness and accuracy were demonstrated with *in vivo* cross-sectional endoscopic OCT images of guinea pig esophagus; Clinical potential was demonstrated by quantitatively comparing the layers' thickness of *in vivo* OCT esophagus images between an eosinophilic esophagitis (EOE) model and a control group.
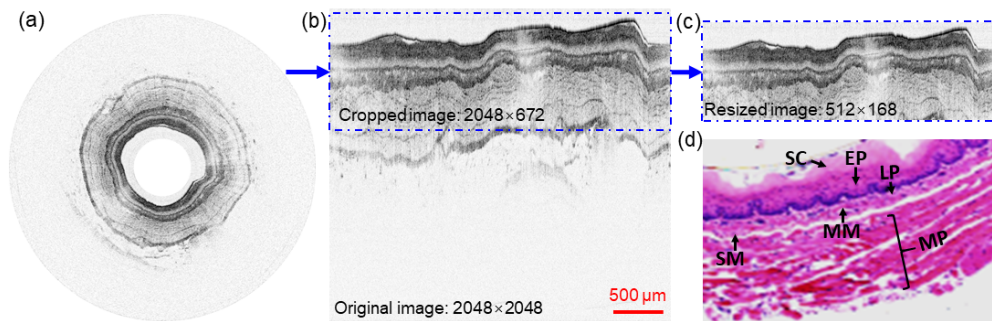

Fig. 1. (a) Representative *in vivo* ultrahigh-resolution circumferential OCT image of guinea pig esophagus. (b) Cut-open (rectangular) OCT image and cropped part. (c) Resized image. (d) Corresponding histology micrograph. SC: stratum corneum, EP: epithelium, LP: lamina propria, MM: muscularis mucosae, SM: submucosa, MP: muscularis propria.

2. Data collection, preparation and control group

Endoscopic OCT esophagus images from five guinea pigs (male, Hilltop, Scottsdale, PA) were collected *in vivo* using an 800 nm endoscopic OCT system (with an OCT endoscope of a 1.3-mm outer diameter). The endoscope was disinfected before deployment into the esophagus. The animals were handled under protocols approved by the Animal Care and Use Committee (ACUC) of the Johns Hopkins University.

Along with a home-built spectral-domain (SD) OCT system [28], the OCT endoscope offered a measured axial resolution of about 2.1 μm in tissue (using a Ti:Sa laser as the low coherence light source with a center wavelength $\lambda_0 = 835nm$ and a spectral bandwidth $\Delta\lambda = 150nm$). With the guinea pig under anesthesia, the OCT endoscope was inserted into the esophagus until reaching the gastro-esophageal junction. 3D imaging was performed by scanning the imaging beam circumferentially during endoscope pullback. Among the five guinea pigs, three were inducted with EOE [33] and the other two served as control. Two hundred cross-sectional images were collected from each guinea pig along the pullback direction. The guinea pigs used in the experiment were at the same age with roughly the same weight. The images used in this manuscript partially overlapped with the ones we previously used to demonstrate our graph-based image processing algorithms [29,34]. Figure 1(a) illustrates a representative cross-sectional OCT image, where the layered structures can be clearly identified by eye (Fig. 1(b)) and correlated well with histology (Fig. 1(d)).

2.1 Data preparation

To facilitate image processing, the circumferential OCT images were converted into a rectangular format, as shown in Fig. 1(b). The rectangular OCT images were then manually segmented based on the normal 6-layer esophagus structure of guinea pig and used to train

and test the deep neural networks. To run the image processing efficiently and avoid potential memory overflow, each rectangular image (2048 × 2048 pixels, lateral × axial) was cropped along depth to keep only the regions of tissues, resulting in a final image size of 2048 × 672 pixels (lateral × axial). The cropped images were further resized down to 512 × 168 pixels by binning every 4 pixels along both the lateral and axial directions (see Fig. 1(c)). The resized images were then augmented by horizontal flipping, spatial translations, and cropping [22]. We trained the networks with 235 OCT images from two guinea pigs with EOE, one control, and their corresponding manual segmentation layer maps (ground truth), among which 215 images served as training data set and the rest 20 images for validation. We tested the trained networks with 40 images and their corresponding ground truth, among which 20 images were from the third EOE guinea pig and 20 images from the second control. It should be mentioned that there was no overlap between the training and testing data sets.
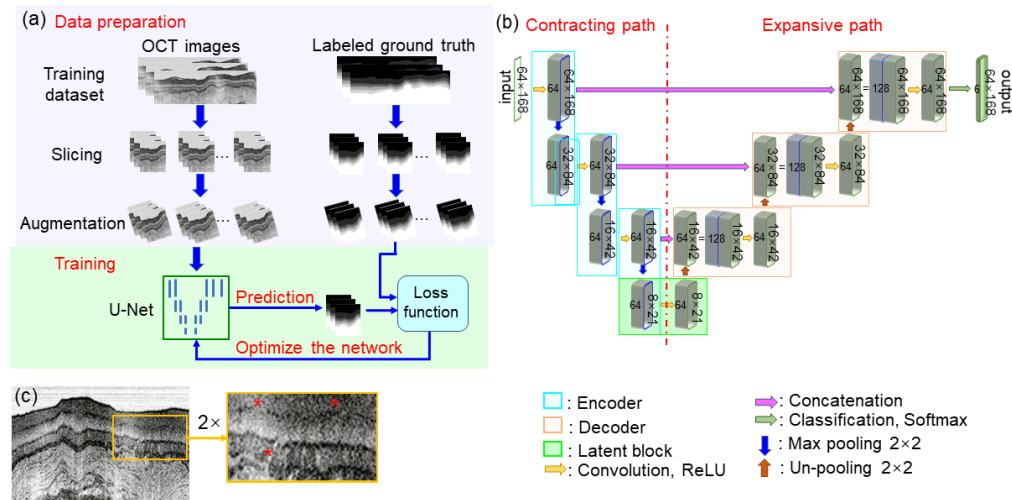

Fig. 2. (a) Schematic of the Net training process. (b) Schematic of a single U-net, which contains 64 kernels for preserving the image features [35]. Rectified Linear Unit (ReLU) induces nonlinearity for efficient training [36]. Max pooling reduces feature maps by a factor of 2 along each dimension [37]. Concatenation helps increase spatial resolution and training stability [38]. (c) Original OCT image with a low image contrast region (see the zoomed-in region indicated by "*").

3. Methods

General description: Our parallel-trained deep neural networks contained three U-Nets [19]. Figure 2(a) shows the training procedure of one representative U-Net. The images and corresponding ground truth in the training data set were first divided along the lateral direction into eight non-overlapping sets of smaller images (termed slices) and then spatial-augmented as the input for the U-Net. The net parameters were initialized randomly, following a normal distribution [39]. The output of the U-Net was the prediction of the esophageal layers. The prediction was compared with the corresponding manual segmentation with a selected loss function. The output of the loss function was used to update the U-Net parameters. The training process was repeated until the loss function reached its minimum. The trained net was then used for automated image segmentation.

3.1 U-net structure

U-Net is fast and precise for medical image segmentation [19,40]. The schematic of our U-net is shown in Fig. 2(b). The network consisted of a contracting path (on the left side of the net), an expansion path (on the right), and a classification layer. There were three encoder blocks (indicated by the cyan color boxes) in the contracting path and three decoder blocks

(indicated by the brown color boxes) in the expansion path. The encoder blocks were used for learning the contextual feature hierarchy and the decoder blocks for semantic segmentation. The decoder blocks were concatenated with the corresponding encoder blocks. The classification block used a convolutional layer with 1×1 kernels to narrow down the feature maps to seven classes, and in our case, the seven classes represent the six esophageal layers and the background. Finally a softmax layer was used to estimate the probability of each pixel belonging to any of the seven classes [41].

Each resized OCT image (512×168 pixels) was laterally divided into eight slices (each of 64×168 pixels) and then spatial augmented. The augmented images were fed into the encoder blocks. In the first encoder block, the input was convolved with 64 kernels to generate $(64 \times 168) \times 64$ feature maps. The feature maps were normalized, activated with ReLU, and sent through a max pooling step for down-sampling by a factor of 2 along both the lateral and depth directions; this process yielded $(32 \times 84) \times 64$ feature maps. The encoding process repeated three times. At the bottom of the U-Net, a latent block served as a transition from the encoder blocks to the decoder blocks, where the $(8 \times 21) \times 64$ feature maps were not down-sampled but instead went through an un-pooling process for up-sampling by a factor of 2 along each dimension, which yielded $(16 \times 42) \times 64$ feature maps. The $(16 \times 42) \times 64$ feature maps were then concatenated with the output from the third encoder block. This concatenated map was convolved, normalized, and activated by ReLU in the remainder of the decoding process. The decoding process repeated twice, and the output of the final decoder block was then sent to the classification block.

3.2 Net parameters update

During the training process, the randomly initialized net parameters were updated for layer prediction by minimizing a loss function. We selected weighted multi-class dice loss function to evaluate the difference between the prediction and manual segmentation [42]. The weighted multi-class dice loss function performed well to compensate for class-imbalance and encouraged kernels that were discriminative towards layer transitions [22,43]. During the training process, we optimized the loss function with an additional Frobenius norm term for regularization [44]. The final loss function is shown below:

$$J = 1 - \frac{2\sum_x p_l(x)g_l(x)}{\sum_x p_l^2(x) + \sum_x g_l^2(x)} + \left\|W\right\|^2 \tag{1}$$

where x is the pixel index, $g_l(x)$ is the true label at pixel x, $p_l(x)$ is the estimated probability for pixel x to belong to class l (there were seven classes in our case), W is the weight for the kernels, and $\left\|W\right\|^2$ is the Frobenius norm term. The output of the loss function layer was used to update the U-Net parameters by trying to reach its minimum. The loss function was minimized by stochastic gradient decent (SGD) with a momentum of 0.9 and an adaptive learning rate during the optimization process [35,45]. The final U-Net parameters were then used for automated image segmentation.

3.3 Parallel training

It has been shown that U-Net can predict retinal layers after being trained by hundreds (at least) of labeled OCT images [23]. The labeled images were either taken from public repositories [22,23,46] or labeled with a well-established software [47]. Labelling endoscopic OCT images is more difficult because of the following factors: (1) any dramatic variation of fine structures on the endoscopic images, (2) geometric complexity induced by tissue folding, or (3) low contrast regions due to sublayers and in-layer fine structures. Due to the first factor, a universally labeled data set or a well-established labeling software don't exist. The last two

factors increase labeling difficulty and inaccuracy. Furthermore, the network also requires a larger training data set to deal with geometric complexity and low image contrast. All these factors result in an elevated cost in data set preparation. Therefore, an effective neural network feasible with a limited training data set is highly desirable.
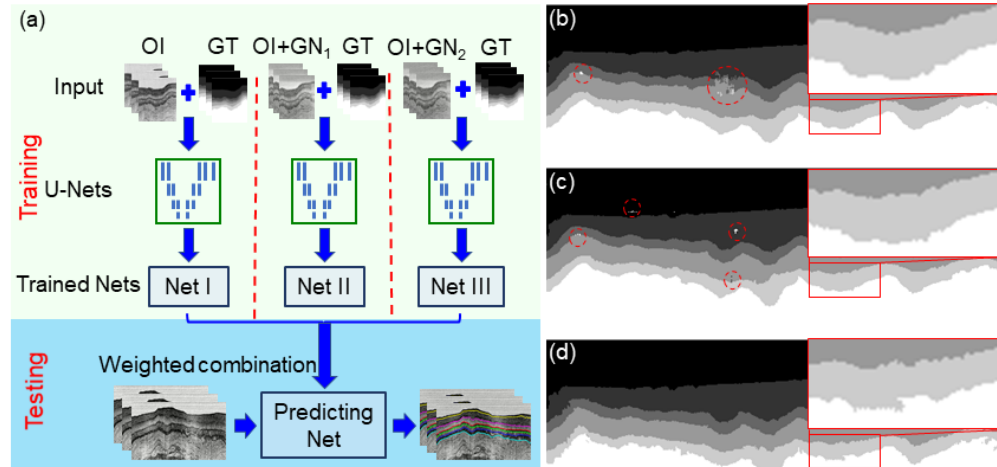

Fig. 3. (a) Flow chart of our parallel training scheme. Three U-Nets (Net I, II, and III) were trained seperately and then combined for layer prediction. Net I was trained by the original images (OI) and the corresponding ground truth (GT); Net II was trained by the original images added with zero-mean Gaussian noise ($\sigma = 1$, GN_1) and the corresponding ground truth; Net III was trained by the original images added with zero-mean Gaussian noise ($\sigma = 2$, GN_2) and the corresponding ground truth. (b) Feature map predicted by the trained Net I. (c) Feature map predicted by the trained Net II. (d) Feature map predicted by the trained Net III. Dashed circles show the regions with layer topology disorders, which decrease from (b) to (d). Zoomed-in boxes show the predictions of layer boundaries, which become noisier from (b) to (d).

Spatial augmentation such as horizontal flipping, translating, and cropping can enlarge the training data set and has served as a standard step in our training stage [48]. However, when we trained a single U-Net (i.e., Net I in Fig. 3(a)) with our limited original training data set, which contained 215 labeled endoscopic OCT images, the prediction of the trained U-Net exhibited layer topology disorders (see the red circle regions in Fig. 3(b)). We also trained two more networks separately by the original training data set added with different levels of zero-mean Gaussian noises (Net II and Net III). Each network was forced to learn topologically correct feature maps within a given noise regime. We noticed that topology disorder decreased when the noise distribution broadened (as shown in the red circle regions in Figs. 3(b-d)), while the predicted boundaries became noisier (as shown in the zoomed-in regions in Figs. 3(b-d)). The trade-off between topology disorder and predicted layer boundary accuracy suggested that a combination of those U-Nets might help achieve a good pixel accuracy and at the same time maintain good shape priors. We then tried different combinations of the U-Nets and compared the outcomes. When any two of the above three networks (Net I, II, and III) were combined, we found topology disorders still existed. When all three networks were combined, topology disorders disappeared. We also noticed that the performance didn't exhibit obvious improvement when more than three networks were combined (i.e., the above three networks plus additional networks trained by the original data set added with broader zero-mean Gaussian noises). Considering the computational time increased almost linearly with the number of networks involved, we adopted the combination of three parallel-trained networks (Net I, II, and III). The combining weights, [0.5, 0.3, 0,2], were selected by minimizing the total layer boundary prediction error for the validation data set (see detailed description of prediction error in Section 4.1).
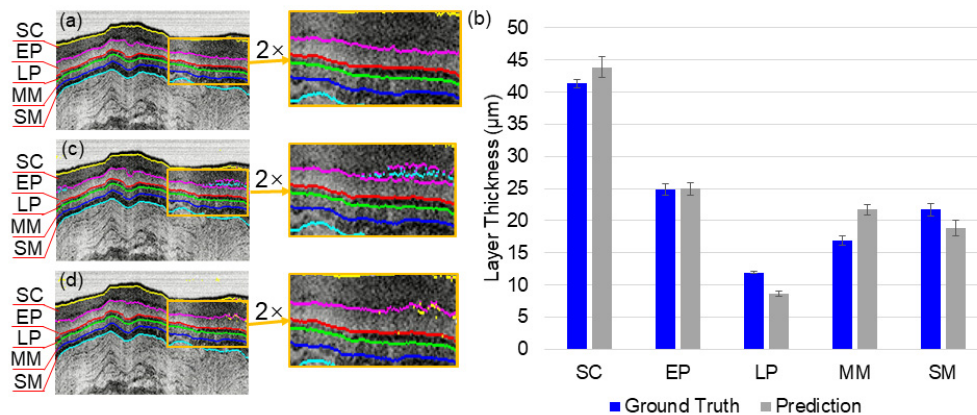
Fig. 4. (a) Representative layer segmentation by the parallel-trained U-Nets (Net I, Net II and Net III) The empirical combination weights are [0.5, 0.3, 0.2]. (b) Layer thickness comparison between the ground truth and the prediction of the parallel-trained U-Nets. Error bars represent the standard deviation of layer thickness for all the images in the testing data set. SC: stratum corneum, EP: epithelium, LP: lamina propria, MM: muscularis mucosae, SM: submucosa. (c) Representative layer segmentation by a single U-Net trained by the original training data set. (d) Representative layer segmentation by a single U-Net trained by a noise-augmented data set which was the combination of the original training data set, the original training data set added with the first Gaussian noises ($\sigma = 1$), and the original training data set added with the second Gaussian noises ($\sigma = 2$).

4. Results

4.1 Evaluation of the parallel training scheme

Figure 4(a) shows representative layer segmentation by the parallel-trained three U-Nets on the testing data set. The result is free from layer topology disorder and the predicted boundaries are smooth as shown in the zoomed-in region of Fig. 4(a). Figure 4(b) illustrates the prediction accuracy in terms of the difference of layers' thickness between prediction and the ground truth. The averaged relative error of prediction (normalized by the ground truth layer thickness) is about 6.0% for all layers. This error might be influenced by non-ideal ground truth. The origin of non-ideal ground truth can be multifaceted. For example, OCT signal saturation at the tissue surface would make it challenging to accurately determine the top boundary of the SC layer, which could result in the large difference between the predicted and ground truth SC layer thickness as seen in Fig. 4(b). In addition, any weak contrast between layers might generate a bias in determining the boundaries of the LP and MM layers, which could lead to the large difference in the predicted and ground truth thickness for the LP and MM layers shown in Fig. 4(b). Error bars in Fig. 4(b) represent the layer thickness variance for each layer for all the images in the testing data set. The absolute boundary error was calculated as the maximum difference of each boundary between the prediction and ground truth. The boundary error averaged over all the boundaries was about 1.4 μm.

In comparison, we also trained a single U-Net with (1) only the original training data set and (2) the noise-augmented data set which was the combination of the original training data set, the original training data set added with the first Gaussian noises ($\sigma = 1$), and the original training data set added with the second Gaussian noises ($\sigma = 2$). Figure 4(c) shows representative layer segmentation by a signal U-Net trained by original data set. Pronounced layer topology disorders (see the boxed region and its zoomed-in view in Fig. 4(c)) occurred in the low contrast area. The averaged error rate of prediction was about 7.0%. Figure 4(d) shows representative layer segmentation by a single U-Net trained by the noise-augmented data set. Layer topology disorders became less pronounced but still visible. The averaged

error rate of prediction was about 6.5%. In comparison with Fig. 4(a), the parallel-trained U-Nets not only reduced layer topology disorders, but also improved segmentation accuracy.

4.2 Differentiation of esophageal layers thickness between the EOE and normal guinea pig models

It has been shown that EOE severity level positively correlated with the thickness of the superficial layers (i.e., the stratum corneum and the epithelium) [49]. To further explore the potential clinical utility of the above parallel-trained U-Nets method, we applied it to differentiating layer thickness of the EOE model and the control using the testing data set.
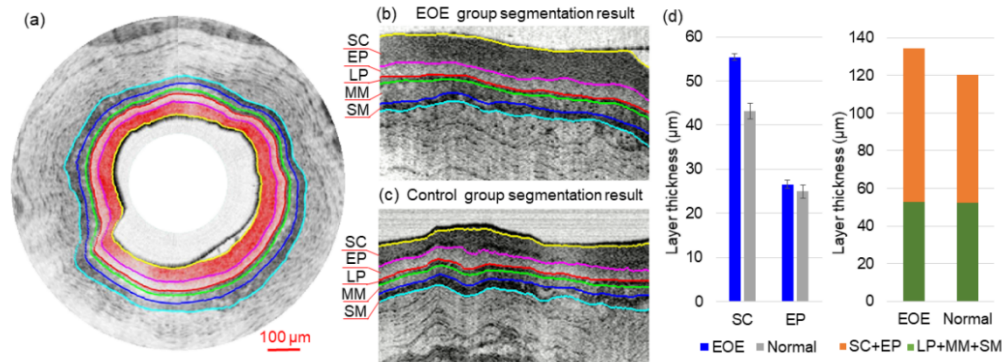

Fig. 5. Layer segmentation and statistics analysis of EOE esophagus model and normal control. (a) Layer segmentations with color-coded SC and EP layers. (b) EOE group layer segmentation result in detail (c) Control group layer segmentation result in detail. (d) Layer thickness comparison between the EOE model and the control group. SC: stratum corneum, EP: epithelium, LP: lamina propria, MM: muscularis mucosae, SM: submucosa, MP: muscularis propria.

Figure 5(a) shows a representative OCT image overlaid with the segmented boundaries where the top-two layers (SC and EP) are color coded. We found that the parallel-trained U-Nets were able to clearly identify and segment the top five layers, as further shown in the zoomed-in regions for the EOE model (Fig. 5(b)) and the control (Fig. 5(c)). One unique value of segmentation is layer thickness quantification. Figure 5(d) shows the layers' thicknesses and comparison between the EOE model and the control. The sum thickness of the top-five layers (averaged over the testing data set) was about 134 µm for the EOE group and 122 µm for the control group. When looking into individual layer, we noticed that (1) the sum thickness of the LP, SM and MM layers remained nearly unchanged, and (2) the SC and EP layers, particularly the SC layer, thickened in the EOE model, with a two-layer sum thickness of 82 µm, which was about 17% thicker than the two-layer sum thickness (70 µm) of the control group. This overall trend was clear and consistent across all images from the testing data set.

5. Discussion

In this paper, we demonstrated that parallel-trained U-Nets can robustly segment the layers in endoscopic OCT images with reduced layer topology disorders. The topology disorders appeared due to the limited training data set, geometric complexity, and low contrast. By combining the U-Nets trained separately with varying levels of Gaussian noise, the layer topology disorders in the prediction were reduced. The proposed scheme demonstrated superb performance mainly due to two reasons: (1) added Gaussian noises prevent overfitting when the original training data set is limited [50]; and (2) separate training of different networks effectively enforces each U-Net to learn shape priors for a given noise regime [51]. In this paper, two zero-mean Gaussian noises of different variances were used for training data set

augmentation. This noise model might not be ideal, as the exact noise model for OCT images of biological tissue is very complex [56]. Nonetheless, this two-variance Gaussian model worked well and was computationally efficient for the parallel-trained deep neural networks scheme. The results show that our prediction accuracy is comparable to the latest published OCT segmentation methods based on deep learning, in which the networks were trained by much larger training data sets with less geometric complexity in the images [52–55]. In the current work, our training and testing data sets focused on OCT images with layered esophageal structures from guinea pig. For endoscopic OCT images collected from other disease models or human subjects with disrupted layer structures such as Barrett's esophagus, the networks need to be re-trained and tested with relevant images.

We also investigated the computational cost for our new method. The computational time for analyzing one image of 2048 × 672 pixels was about 0.6 s on a Windows computer with a 4-core, 4.2-GHz CPU and a GPU with 4 GB memory with the codes implemented in MATLAB. The speed was about ten times faster than our previously reported graph-based methods [29]. The speed is expected to improve dramatically with a hardware upgrade and implementation in C + + , which would be very attractive for future real-time layer segmentation and tracking in various clinical applications.

Funding

Disclosures

The authors declare that there are no conflicts of interest related to this article. The code will be available by emailing request to jhu.bme.bit@gmail.com.

References

1. A. K. S. Jardine, D. M. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," Mech. Syst. Signal Process. **20**(7), 1483–1510 (2006).
2. C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Tech. Rep. (2005).
3. W. Hatta, K. Uno, T. Koike, S. Yokosawa, K. Iijima, A. Imatani, and T. Shimosegawa, "Optical coherence tomography for the staging of tumor infiltration in superficial esophageal squamous cell carcinoma," Gastrointest. Endosc. **71**(6), 899–906 (2010).
4. I.-K. Jang, B. E. Bouma, D.-H. Kang, S.-J. Park, S.-W. Park, K.-B. Seung, K.-B. Choi, M. Shishkov, K. Schlendorf, E. Pomerantsev, S. L. Houser, H. T. Aretz, and G. J. Tearney, "Visualization of coronary atherosclerotic plaques in patients using optical coherence tomography: comparison with intravascular ultrasound," J. Am. Coll. Cardiol. **39**(4), 604–609 (2002).
5. M. J. Gora, M. J. Suter, G. J. Tearney, and X. Li, "Endoscopic optical coherence tomography: technologies and clinical applications [Invited]," Biomed. Opt. Express **8**(5), 2405–2444 (2017).
6. D. C. Adler, C. Zhou, T. H. Tsai, J. Schmitt, Q. Huang, H. Mashimo, and J. G. Fujimoto, "Three-dimensional endomicroscopy of the human colon using optical coherence tomography," Opt. Express **17**(2), 784–796 (2009).
7. S. Tsantis, G. C. Kagadis, K. Katsanos, D. Karnabatidis, G. Bourantas, and G. C. Nikiforidis, "Automatic vessel lumen segmentation and stent strut detection in intravascular optical coherence tomography," Med. Phys. **39**(1), 503–513 (2011).
8. G. J. Ughi, T. Adriaenssens, K. Onsea, P. Kayaert, C. Dubois, P. Sinnaeve, M. Coosemans, W. Desmet, and J. D'hooge, "Automatic segmentation of in-vivo intra-coronary optical coherence tomography images to assess stent strut apposition and coverage," Int. J. Cardiovasc. Imaging **28**(2), 229–241 (2012).
9. H. Lu, M. Gargesha, Z. Wang, D. Chamie, G. F. Attizzani, T. Kanaya, S. Ray, M. A. Costa, A. M. Rollins, H. G. Bezerra, and D. L. Wilson, "Automatic stent detection in intravascular OCT images using bagged decision trees," Biomed. Opt. Express **3**(11), 2809–2824 (2012).
10. Z. Wang, D. Chamie, H. G. Bezerra, H. Yamamoto, J. Kanovsky, D. L. Wilson, M. A. Costa, and A. M. Rollins, "Volumetric quantification of fibrous caps using intravascular optical coherence tomography," Biomed. Opt. Express **3**(6), 1413–1426 (2012).
11. G. J. Ughi, M. J. Gora, A.-F. Swager, A. Soomro, C. Grant, A. Tiernan, M. Rosenberg, J. S. Sauk, N. S. Nishioka, and G. J. Tearney, "Automated segmentation and characterization of esophageal wall in vivo by tethered capsule optical coherence tomography endomicroscopy," Biomed. Opt. Express **7**(2), 409–419 (2016).

12. M. Mujat, R. Chan, B. Cense, B. Park, C. Joo, T. Akkin, T. Chen, and J. de Boer, "Retinal nerve fiber layer thickness map determined from optical coherence tomography images," Opt. Express **13**(23), 9480–9491 (2005).
13. A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi, "Intra-retinal layer segmentation in optical coherence tomography images," Opt. Express **17**(26), 23719–23728 (2009).
14. D. Cabrera Fernández, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," Opt. Express **13**(25), 10200–10216 (2005).
15. A. M. Bagci, M. Shahidi, R. Ansari, M. Blair, N. P. Blair, and R. Zelkha, "Thickness profiles of retinal layers by optical coherence tomography image segmentation," Am. J. Ophthalmol. **146**(5), 679–687 (2008).
16. Q. Yang, C. A. Reisman, K. Chan, R. Ramachandran, A. Raza, and D. C. Hood, "Automated segmentation of outer retinal layers in macular OCT images of patients with retinitis pigmentosa," Biomed. Opt. Express **2**(9), 2493–2503 (2011).
17. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," Opt. Express **18**(18), 19413–19428 (2010).
18. S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," Biomed. Opt. Express **6**(4), 1172–1194 (2015).
19. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lect. Notes Comput. Sci. **9351**, 234–241 (2015).
20. K. J. Han and Y. H. Lee, "Optical coherence tomography automated layer segmentation of macula after retinal detachment repair," PLoS One **13**(5), e0197058 (2018).
21. F. G. Venhuizen, B. van Ginneken, B. Liefers, M. J. J. P. van Grinsven, S. Fauser, C. Hoyng, T. Theelen, and C. I. Sánchez, "Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks," Biomed. Opt. Express **8**(7), 3292–3316 (2017).
22. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger, and N. Navab, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," Biomed. Opt. Express **8**(8), 3627–3642 (2017).
23. L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," Biomed. Opt. Express **8**(5), 2732–2744 (2017).
24. A. Montuoro, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, and H. Bogunović, "Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context," Biomed. Opt. Express **8**(3), 1874–1888 (2017).
25. Y. He, A. Carass, Y. Yun, C. Zhao, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Towards topological correct segmentation of macular oct from cascaded fcns," in *Fetal, Infant and Ophthalmic Medical Image Analysis* (Springer, 2017), pp. 202–209.
26. A. Abdolmanafi, L. Duong, N. Dahdah, and F. Cheriet, "Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography," Biomed. Opt. Express **8**(2), 1203–1220 (2017).
27. S. K. Devalla, K. S. Chin, J.-M. Mari, T. A. Tun, N. G. Strouthidis, T. Aung, A. H. Thiéry, and M. J. A. Girard, "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head," Invest. Ophthalmol. Vis. Sci. **59**(1), 63–74 (2018).
28. J. Xi, A. Zhang, Z. Liu, W. Liang, L. Y. Lin, S. Yu, and X. Li, "Diffractive catheter for ultrahigh-resolution spectral-domain volumetric OCT imaging," Opt. Lett. **39**(7), 2016–2019 (2014).
29. J. Zhang, W. Yuan, W. Liang, S. Yu, Y. Liang, Z. Xu, Y. Wei, and X. Li, "Automatic and robust segmentation of endoscopic OCT images and optical staining," Biomed. Opt. Express **8**(5), 2697–2708 (2017).
30. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Med. Image Anal. **42**, 60–88 (2017).
31. H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016).
32. P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning for multi-task medical image segmentation in multiple modalities," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2016), 478–486.
33. Z. Liu, Y. Hu, X. Yu, J. Xi, X. Fan, C. M. Tse, A. C. Myers, P. J. Pasricha, X. Li, and S. Yu, "Allergen challenge sensitizes TRPA1 in vagal sensory neurons and afferent C-fiber subtypes in guinea pig esophagus," Am. J. Physiol. Gastrointest. Liver Physiol. **308**(6), G482–G488 (2015).
34. M. Gan, C. Wang, T. Yang, N. Yang, M. Zhang, W. Yuan, X. Li, and L. Wang, "Robust layer segmentation of esophageal OCT images based on graph search using edge-enhanced weights," Biomed. Opt. Express **9**(9), 4481–4495 (2018).
35. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Commun. ACM **60**(6), 84–90 (2017).
36. R. Livni, S. Shalev-Shwartz, and O. Shamir, "On the computational efficiency of training neural networks," in *Advances in Neural Information Processing Systems*, 2014), 855–863.

37. M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Advances in neural information processing systems*, 2015), 2017–2025.
38. A. E. Orhan and X. Pitkow, "Skip connections eliminate singularities," arXiv preprint arXiv:1701.09175 (2017).
39. Y. LeCun, L. Bottou, G. B. Orr, and K. R. Muller, "Efficient backprop," Neural Networks: Tricks of the Trade **1524**, 9–50 (1998).
40. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2016), 424–432.
41. Y. Tang, "Deep learning using linear support vector machines," arXiv preprint arXiv:1306.0239 (2013).
42. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Springer, 2017), pp. 240–248.
43. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, (IEEE, 2016), 565–571.
44. S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-Based Norm Regularization for Image Restoration with Biomedical Applications," IEEE Trans. Image Process. **21**(3), 983–995 (2012).
45. L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the Trade* (Springer, 2012), pp. 421–436.
46. A. Shah, L. Zhou, M. D. Abràmoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images," Biomed. Opt. Express **9**(9), 4509–4526 (2018).
47. J. Hamwood, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers," Biomed. Opt. Express **9**(7), 3049–3066 (2018).
48. L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621 (2017).
49. S. J. Chien, K. A. Silva, V. E. Kennedy, H. HogenEsch, and J. P. Sundberg, "The pathogenesis of chronic eosinophilic esophagitis in SHARPIN-deficient mice," Exp. Mol. Pathol. **99**(3), 460–467 (2015).
50. C. M. Bishop, "Training with Noise Is Equivalent to Tikhonov Regularization," Neural Comput. **7**(1), 108–116 (1995).
51. A. Romero, M. Drozdzal, A. Erraqabi, S. Jégou, and Y. Bengio, "Image Segmentation by Iterative Inference from Conditional Score Estimation," arXiv preprint arXiv:1705.07450 (2017).
52. J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search," Biomed. Opt. Express **9**(11), 5759–5777 (2018).
53. Y. Guo, A. Camino, M. Zhang, J. Wang, D. Huang, T. Hwang, and Y. Jia, "Automated segmentation of retinal layer boundaries and capillary plexuses in wide-field optical coherence tomographic angiography," Biomed. Opt. Express **9**(9), 4429–4442 (2018).
54. A. Shah, L. Zhou, M. D. Abràmoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images," Biomed. Opt. Express **9**(9), 4509–4526 (2018).
55. Y. He, A. Carass, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Topology guaranteed segmentation of the human retina from OCT using convolutional neural networks," arXiv preprint arXiv:1803.05120 (2018).
56. T. B. Dubose, D. Cunefare, E. Cole, P. Milanfar, J. A. Izatt, and S. Farsiu, "Statistical models of signal and noise and fundamental limits of segmentation accuracy in retinal optical coherence tomography," IEEE Trans. Med. Imaging **37**(9), 1978–1988 (2018).